(51) International Patent Classification 5 : **H04L 12/56**	**A1**	**(11) International Publication Number:** **WO 94/14263**
		(43) International Publication Date: 23 June 1994 (23.06.94)

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PYHÄLAMMI, Seppo [FI/FI]; Sankaritie 4 A 1, FIN-00320 Helsinki (FI). SALOVUORI, Heikki [FI/FI]; Fregattikuja 3, FIN-00850 Helsinki (FI). HARJU, Esa [FI/FI]; Avaruuskatu 4 B 19, FIN-02210 Espoo (FI). HAVANSI, Klaus [FI/FI]; Ristiaallokonkatu 4 F 132, FIN-02320 Espoo (FI). MATKASELKÄ, Jorma [FI/FI]; Nuijatie 4 C 24, FIN-01650 Vantaa (FI).
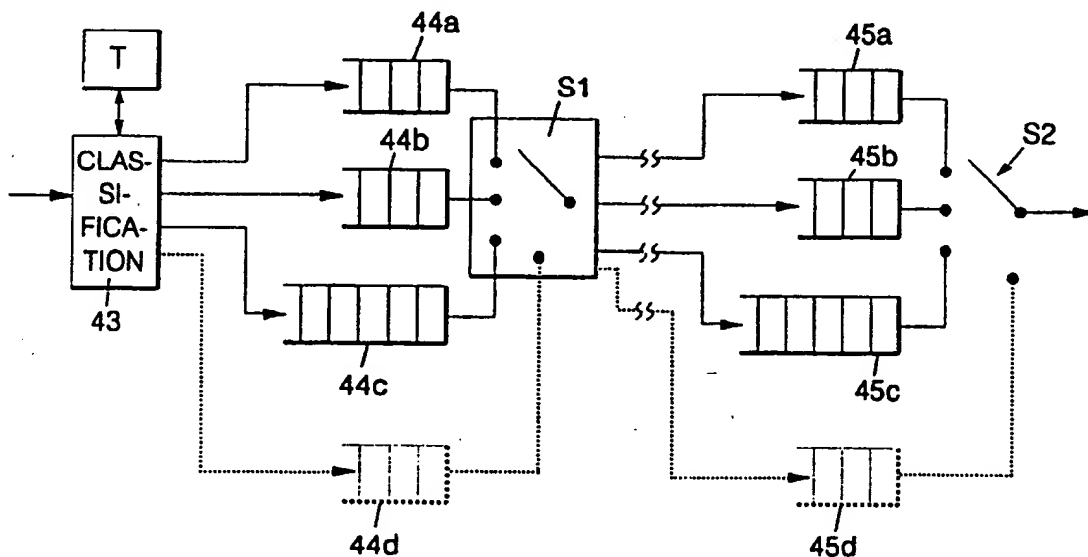
(74) Agent: OY KOLSTER AB; Iso Roobertinkatu 23, P.O. Box 148, FIN-00120 Helsinki (FI).

(54) Title: A METHOD FOR CONGESTION MANAGEMENT IN A FRAME RELAY NETWORK AND A NODE IN A FRAME RELAY NETWORK

(57) Abstract

The invention relates to a method for congestion management in an FR network and to a node in an FR network. In the method, frames (39) are buffered both at the input and at the output boundary of the node, and a virtual channel associated with the frame (39) to be transferred is determined. In order to provide a new, reliable congestion management method allowing rapid responding and capable of offering even the user application behind the congestion service as well as congestion management mechanisms that reduce the level of service as little as possible, network services are classified into different classes (1, 2, 3) according to delay and frame throughput probability; and frames are forwarded in the nodes through service class specific buffers (44a...44d, 45a...45d, 64a...64d).

1

A method for congestion management in a frame relay
network and a node in a frame relay network

5   The invention relates to a method for congestion
management in a frame relay network, wherein frames
are buffered both at the input and at the output
boundary of a node, and a virtual channel associated
with a frame to be transferred is determined. The
invention further relates to a node in a frame relay
10   network, comprising input buffers at the input
boundary of the node and output buffers at the output
boundary of the node, and means for delivering frames
from the input buffer to a desired output buffer.

Congestion means a situation in which the number
15   of transmission requests exceeds the transmission
capacity at a certain network point (called a bottle-
neck resource) at a specific time. Congestion usually
results in overload conditions, as a result of which
the buffers overflow, for instance, and so packets
20   will be retransmitted either by the network or the
subscriber. The function of congestion management (CM)
is to maintain a balance between the transmission
requests and the transmission capacity so that the
bottle-neck resources operate on an optimal level, and
25   the subscribers are offered service in a way that
assures fairness.

Congestion management can be divided into
congestion avoidance (CA) and congestion recovery
(CR). Congestion avoidance methods aim at preventing
30   the occurrence of congestion in the network by
dynamically adjusting the bandwidth of the subscribers
in accordance with network congestion conditions
and/or by altering the network routing so that part of
the traffic load of the bottle-neck resources is
35   shifted to idle resources. The purpose of recovery

2

methods, in turn, is to restore the operation of the
bottle-neck resources to an optimal level if the
avoidance methods have failed to prevent the
occurrence of congestion.

5   The frame relay (FR) technique is a packet-
switched network technique used for the transmission
of frames of varying length in place of the packet-
switched network connections presently in use. The
protocol (X.25) applied generally in the present

10   packet-switched networks requires plenty of processing
and the transmission equipment is expensive, as a
result of which the speeds also remain low. These
matters are due to the fact that the X.25 standard was
developed when the used transmission connections were

15   still rather prone to transmission errors. The
starting point of the frame relay technique was a
considerably lower transmission line error probab-
ility. It has therefore been possible to discard a
number of unnecessary functions in the frame relay

20   technique, thus making the delivery of frames rapid
and efficient. The Frame Mode Bearer Service is
described generally in the CCITT specification I.233
(Reference 1) and the associated protocol in the
specification Q.922 (Reference 2). Congestion in an FR

25   network and congestion management mechanisms are
described in the CCITT specification I.370 (Reference
3). For a more detailed description of the FR
technique, *An Overview of Frame Relay Technology*,
Datapro Management of Data Communications, McGraw-Hill

30   Incorporated, April 1991, (Reference 4) as well as the
above-mentioned specifications are referred to.
          The recommendations defined in the CCITT
specifications offer a few mechanisms for congestion
notification. For instance, the recommendation I.370

35   shares the network congestion management between the

nodes and the subscribers of the network. According to it, the subscriber should adjust the amount of traffic on the basis of congestion notifications received from the network, and the network, in turn, should do the same on the basis of congestion notifications received from the subscriber. In other words, if the network notifies the subscriber of congestion, the subscriber should reduce its traffic to the network, and *vice versa*. This co-operation would thus make the network a kind of closed system as far as the congestion management is concerned. However, this kind of congestion management is not operative in practice, mainly for two reasons:

- notification mechanisms offered by the specification are too slow to respond to momentary congestion situations; and

- one cannot rely on the subscribers voluntarily reducing traffic even if they receive congestion notifications from the network. In such a case the congestion management system is not, in fact, closed, and congestion will not be relieved.

Another drawback of known methods is that all subscribers are treated equally under congestion conditions, and so it is not possible to give priority to the messages of subscribers wanting better service (i.e. applications requiring better service) with respect to the most important application parameters.

The object of the present invention is to eliminate the above-described disadvantages and provide a new congestion management method for use in the FR network, which method is reliable and capable of rapid responding. The new method enables the subscriber application behind the congestion to be offered the best possible service as well as congestion management mechanisms that reduce the level

4

of service as little as possible. This is achieved by
a method according to the invention, which is
characterized in that network services are classified
into different classes according to delay and frame
5   throughput probability; and frames are forwarded in
the nodes through buffers provided specifically for
each service class. An FR network node according to
the invention, in turn, is characterized in that it
comprises   means   for   storing   service   classes
10   corresponding to virtual channels; and at least one
buffer for each defined service class provided at
least at the output boundary of the node specifically
for each data link connection.

The invention rests on the idea that the
15   services offered in the network are classified accord-
ing to the most important application parameters
(throughput probability and delay) and that the
delivery of frames to be transferred is performed
specifically for each service class. In practice, this
20   means e.g. that the different service classes are
treated in different ways as far as buffering and
congestion management are concerned, which in turn
means mainly that a separate buffer is required for
each service class for the allocation of shared
25   resources within each access connection.

In the following the invention will be described
in greater detail with reference to the examples shown
in the attached drawings, in which

Figure 1 illustrates the classification of
30   services in an FR network according to the invention;

Figure 2 illustrates a typical operating
environment of the method according to the invention;

Figure 3 is a simplified view of a node in the
FR network;

35   Figure 4 illustrates the principal features of

the configuration of a trunk node in the FR network
according to the invention;

Figure 5 illustrates the format of a frame to be
delivered in the FR network;

Figure 6 illustrates the principal features of a
subscriber node in the FR network according to the
invention;

Figure 7 illustrates the transmission of con-
gestion notifications in the network for adjusting
subscriber-specific traffic; and

Figure 8 shows the trunk node of Figure 4 with a
further service class added to the network for
congestion notifications.

The frame relay network supports various appli-
cations requiring services of different kinds. For
example, if one generalizes to some extent, traffic
between local area networks (LAN) can be divided into
two portions: (a) interactive traffic, and (b) data-
file transfer type traffic. In interactive traffic,
relatively short packets are transferred and a rapid
response is expected from the network, i.e. small
delays. Transmission is usually carried out by higher-
order protocols which detect transmission errors and
frame losses and perform a rapid retransmission, if
required.

In the datafile transfer, it is beneficial to
transfer relatively large packets. The time required
for the transfer is not usually critical, provided
that the time-outs of the protocols will not expire.

In addition to the services described above,
there is a need for a service where both delays and
frame throughput probability are optimized. This kind
of service requires more network node resources than
the other services described above, and would, in
practice, also be more expensive than the other

6

services.

On the basis of the above, the services of the FR network can be described as shown in Figure 1/ taking into account the two most important application parameters (frame loss probability and delay). Area 1 represents a service offering interactivity; area 2 represents data transfer service; and area 3 represents a service where both delays and frame throughput probability are optimized. With a low traffic rate in the FR network, the operation takes place near the origin, and the different services do not differ to any greater degree from each other. The network is able to forward all traffic, and there occur no packet losses due to congestion. However, momentary network congestion cannot be avoided, and network nodes should respond in such situations as efficiently as possible. When the traffic increases, the operation shifts farther away from the origin, and differences between the different services become apparent. Service 2 is able to utilize a long buffering time to achieve a good multiplexing result. This causes a delay to occur but this is allowable for this particular service. Service 1 is able to utilize the fact that delays must not be too long. As a last resort, frames can be discarded as congestion would anyway increase their transit delay excessively. Service 3 aims at both a short delay and a high throughput probability, and so it requires more capacity in the nodes.

Figure 2/ shows an FR network offering public network services, that is, a frame relay network 12 interconnecting local area networks 11 of different offices A...E of a single corporation or a plurality of corporations. The local area network 11 of each office has access to the FR service via a local area

network bridge 13 and a data link connection indicated
with the references 14a...14e, respectively. The con-
nection between an FR subscriber A...E and an FR
network node is known *per se*, wherefore it will not be
described more closely herein. (More detailed informa-
tion about local area networks and bridges used in
their interconnection can be found e.g. in an article
by Michael Grimshaw *LAN Interconnections Technology*,
Telecommunications, February 1991, and in *Lähiverkko-
opas*, Leena Jaakonmäki, Suomen ATK-kustannus Oy, 1991,
which are incorporated herein by reference.)

Figure 3 is a simplified view of the known
structure of a node N in the FR network 12. An FR
frame from a subscriber is received in an input buffer
15a, from which it is connected to a centralized
router 16 routing it to an appropriate output buffer
15b or 15c, which may be either an output buffer 15b
connected to another subscriber connection or an
output buffer 15c connected to an FR trunk connection
(which is an internodal data link connection in the FR
network).

A typical feature of the node structure shown in
Figure 3 is that the same buffer is used for all
frames, assuming that they are routed to the same
physical connection. According to the present inven-
tion, on the contrary, buffers corresponding to the
above-described service classes are provided at the
output boundary of all network nodes and at the input
boundary having trunk connections. Figure 4 illus-
trates this kind of solution at a trunk node in the
network. The node receives FR frames originally
assembled in the bridges 13 of the subscriber
connections (Figure 2). The frame of the subscriber
LAN 11 is inserted into the information field of the
FR frame in the bridge 13 (with the exception of

8

timing bits and other similar bits). Figure 5
illustrates the insertion of a LAN frame 38 into the
information field of an FR frame 39. It also shows a
typical FR network frame format where the address
5   field preceding the information field comprises two
octets (bits 1 to 8). The bits 3 to 8 of the first
octet and the bit 5 to 8 of the second octet form a
data link connection identifier DLCI, which indicates
to the node e.g. the virtual connection and virtual
10   channel to which a particular frame belongs. The
virtual channels are distinguished from each other by
means of the data link connection identifier. The data
link connection identifier, however, is unambiguous
only over a single virtual channel, and it may change
15   in the node on transition to the next virtual channel.
In the invention, the bit 2 of the second address
field octet, called a DE bit (Discard Eligibility
Indicator), is also significant. In accordance with
the CCITT recommendation, it is allowable to discard a
20   frame e.g. under congestion conditions if the DE bit
of the frame has been set to one. For instance, in the
service class 1, which tolerates a higher frame loss
probability than the other service classes, such
frames are discarded more readily than in the other
25   classes as the congestion would anyway increase the
transit delay of the frames excessively. As the other
bits in the FR frame are not relevant to the present
invention, they will not be described more closely
herein. References 2 and 4 mentioned above are
30   referred to for a more detailed description.

The FR frame 39 of the format described above is
received at the node by a special classification unit
43, each data link connection having its own classi-
fication unit, which reads the identifier DLCI from
35   the address field of the frame and selects the service

classification

class (1, 2 or 3) corresponding to the virtual channel
indicated by the identifier. The virtual channels and
the respective service classes may be stored e.g. in a
table T. (The virtual channel means a virtual con-
5   nection portion having the length of one transmission
interval, while the virtual connection is the actual
packet-switched end-to-end FR connection.) After
having completed the classification by means of the
table T, the classification unit 43 applies each frame   *input buffer*
10   to an input buffer 44a, 44b or 44c corresponding to
the specified service class of the frame. Each inbound
data link connection thus comprises three input
buffers, one for each service class. Each data link
connection has a specific selector S1 which reads the
15   frames from the service class specific buffers and
forwards them within the node. On the output side of   *— output buffer*
the trunk node, each frame 39 is applied to one of
three service class specific output buffers 45a, 45b
or 45c of a desired data link connection, the buffer
20   being selected according to the service class
specified on the input side of the node. A selector S2
then reads the frame further to a trunk connection.
Alternatively, classification units may be provided
separately for each transmission connection on the
25   output side of the node, in which case the class data
need not be transferred within the node.

In the service class 2 (buffers 44c and 45c in
Figure 4), a lot of traffic can be buffered, thus
making the transmission more reliable at the expense
30   of delays. In the service class 1 (buffers 44a and
45a), the size of the buffers should be small (smaller
than in class 2) and the same applies to the service
class 3 (buffers 44b and 45b). (The size of the buffer
is determined on the basis of the maximum delay the
35   network is allowed to cause.) The frames thus pass

through the network nodes with a small delay depending on the service.

The selectors S1 and S2 read the buffers in proportion to the traffic allocated to them, so that the fairness principle will be met, or by applying some other suitable principle so that the agreed service will be provided. As the service class 3 further involves a promise of a lower loss probability under congestion conditions, the buffers 44b and 45b have to be read with a higher priority than the other buffers (for example, more frequently than the other buffers). In addition, when the fill rate of each service class specific buffer exceeds a predetermined limit, the discarding of preferentially frames having their DE bit set (to one) can be started.

Figure 6 illustrates a subscriber node N according to the invention (i.e. a node to which subscribers are connected). Subscriber connections 14a, 14b, etc. are connected to an identification unit 61 which receives FR frames which are of the type described above and have been assembled in the bridges 13 (Figure 2). The identification unit 61 reads the identifier DLCI from the address field of the frame and forwards the frame to an input buffer $61_1...62_n$ corresponding to the virtual connection indicated by the identifier. Each data link connection has a specific selector S3 which selects frames from the input buffers of each virtual channel and forwards them within the node.

On the output side of the node, frames are applied to the above-described classification unit 43, each data link connection having its own classification unit, which reads the identifier DLCI from the address field of the frame and selects from table T the service class corresponding to the virtual channel

11

indicated by the identifier. After having completed the classification, the classification unit 43 applies each frame to an output buffer 64a, 64b or 64c corresponding to the frame's service class. Each outbound

5   data link connection thus has three output buffers, one for each service class. The selector S2 selects the frames from the service class specific output buffers 64a...64c and forwards them to the data link connection. (The function of the fourth service class

10   buffer drawn by broken lines in the figure will be described more closely below.)

Traffic transmitted by the subscribers over the FR network is thus buffered on the input side of the subscriber node specifically for each virtual

15   connection. Incoming frames 39 are chained dynamically over each virtual connection. Depending on the service class of the virtual connection, the chain length has a predetermined allowable maximum value; this is smaller in service classes 1 and 3 and greater in

20   service class 2. The selector S3 reads the buffers $62_1...62_n$ e.g. in proportion to the amount of traffic allocated to them, thus meeting the fairness principle, or by applying some other suitable principle so as to provide the agreed service.

25   On the output side of the subscriber node, the service class specific buffers are utilized in the same way as in the trunk node.

According to a preferred embodiment of the invention, the service to be provided in each service

30   class can be further improved so that when a network node is congested, one attempts to reduce the amount of traffic over the virtual connections passing through the node already at the source end of the network in the subscriber node of the particular

35   virtual connections. Accordingly, frames will not load

12

the other network resources only to be discarded on reaching the congested node. This mechanism operates in the following way (cf. Figure 7). When the output buffers in a network node, say a node N1, exceed a predetermined fill rate, the node transmits a special congestion notification M in a backward direction to the subscriber nodes (node N2 in the figure); the amount of traffic transmitted to the network over the virtual connections is thus reduced by buffering such traffic into the virtual-channel-specific buffers $62_1 \ldots 62_n$ to such an extent as possible. After a while, if no further congestion notifications are received, the amount of traffic through this particular resource may again be increased. By monitoring traffic specifically for each subscriber by utilizing the congestion notifications, the amount of traffic to a congested resource can be reduced in order to relieve congestion.

In practice, the adjustment of the amount of traffic in the subscriber node is performed entirely (in compliance with the fairness principle) by monitoring the traffic of each subscriber (S) specifically for each virtual connection by means of parameters CIR (Committed Information Rate), B_c (Committed Burst Size) and B_e (Excess Burst Size), and by responding to the above-described congestion notifications M by restricting the transmission of traffic to the network over certain virtual connections in order to relieve congestion. CIR represents the data transmission rate guaranteed by the network under normal conditions; B_c represents a maximum amount of data the subscriber can transmit over the network within a time slot T_c; and B_e represents the amount of data by which the subscriber can exceed the value B_c within the time slot T_c. These

parameters are described e.g. in Reference 3.

Congestion notifications M should be forwarded to the egress nodes very rapidly for rapid responding under congestion conditions. According to the pre-
5   ferred embodiment of the invention, a fourth service class is dedicated to such congestion notifications, and respective buffers are provided for the service class in the nodes. For the trunk node, this embodiment is shown in Figure 8 where the buffers 44d and
10   45d of the service class of congestion notifications are drawn by broken lines. Correspondingly, a fourth output buffer 64d is incorporated on the output side of the subscriber node (Figure 6).

     The mechanism described above makes the FR
15   network a closed system as far as the congestion management is concerned. When a congestion notifi-cation M is transmitted in a backward direction to the subscriber nodes sufficiently early before the node is severely congested, the network and network traffic
20   can be "tuned" flexibly in this way to achieve an optimal throughput capacity. The service class specific buffers ensure that each service class is treated in a way best suited for its characteristic features even under congestion conditions.

25   Even though the invention has been described above with reference to the examples of the attached drawings, it is obvious that the invention is not restricted to it but it may be modified within the inventive idea disclosed above and in the attached
30   claims. For example, the internal structural details of the nodes may vary in many ways even though the service classes and respective buffers are implemented in accordance with the inventive idea. It is further to be noted that the above-described way is used in
35   the nodes to forward only frames the integrity of

which has been checked and in which no errors have been detected and the DLCI has been defined for the link in question. If errors are detected or it is found out that the DLCI has not been defined for the
5   link in question, the frame is discarded. As these stages, however, do not fall within the idea of the invention, they have not been described above.

15

Claims:

1. Method for congestion management in a frame relay network, wherein frames (39) are buffered both at the input and at the output boundary of a node, and a virtual channel associated with the frame (39) to be transferred is determined, c h a r a c t e r i z e d in that

- network services are classified into different classes (1, 2, 3) according to delay and frame through-put probability; and

- frames are forwarded in the nodes through buffers (44a...44d, 45a...45d, 64a...64d) specific for each service class.

2. Method according to claim 1, c h a r a c-t e r i z e d in that the network services are divided into three main service classes so that the first service class (1) offers interactive service where delays are kept short; the second service class (2) offers a low frame loss probability while delays are less critical; and the third service class (3) offers both a short delay and a low frame loss probability.

3. Method according to claim 1, c h a r a c-t e r i z e d in that the buffers of the third service class are read more frequently than those of the other classes.

4. Method according to claim 1, c h a r a c-t e r i z e d in that frames are buffered into buffers ($62_1...62_n$) specific for each virtual channel at the input boundary of the subscriber node.

5. Method according to claim 4, c h a r a c-t e r i z e d in that a further service class is dedicated to congestion notifications (M) transmitted from each congested resource to the egress node of a

16

particular virtual connection, traffic from a subscriber (S) being adjusted in the egress node by buffering frames into the virtual channel specific buffers ($62_1 \ldots 62_n$).

5   6. Node in a frame relay network, comprising input buffers at the input boundary of the node and output buffers at the output boundary of the node, and means (16) for forwarding frames from the input buffer to a desired output buffer, c h a r a c t e r i z e d

10   in that it comprises

- means (T) for storing service classes (1, 2, 3) corresponding to virtual channels; and

- at least one buffer (45a, 64a; 45b, 64b; 45c, 65c) for each defined service class provided at least

15   at the output boundary of the node specifically for each data link connection.

7. Subscriber node according to claim 6, c h a r a c t e r i z e d in that buffers ($62_1 \ldots 62_n$) specific for each virtual channel are provided at the

20   input boundary of the node.

8. Trunk node according to claim 6, c h a r a c- t e r i z e d in that buffers (44a...44d) specific for each service class are also provided at the input boundary of the node.

25   9. Node according to claim 6, c h a r a c- t e r i z e d in that the buffers of the service classes offering a delay shorter than that offered by the other service classes are shorter than the buffers of the other service classes.
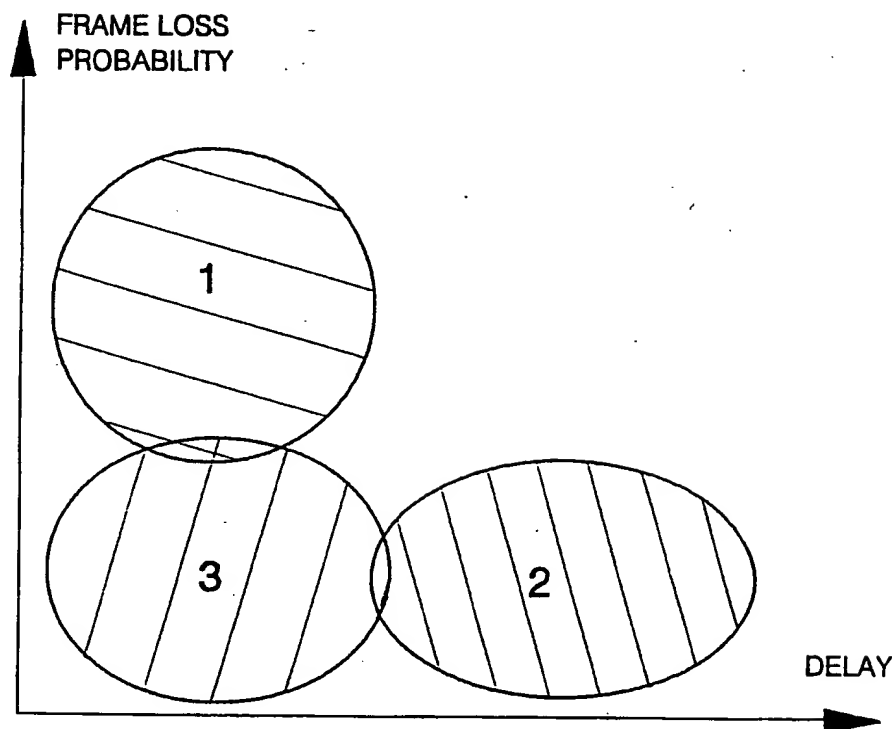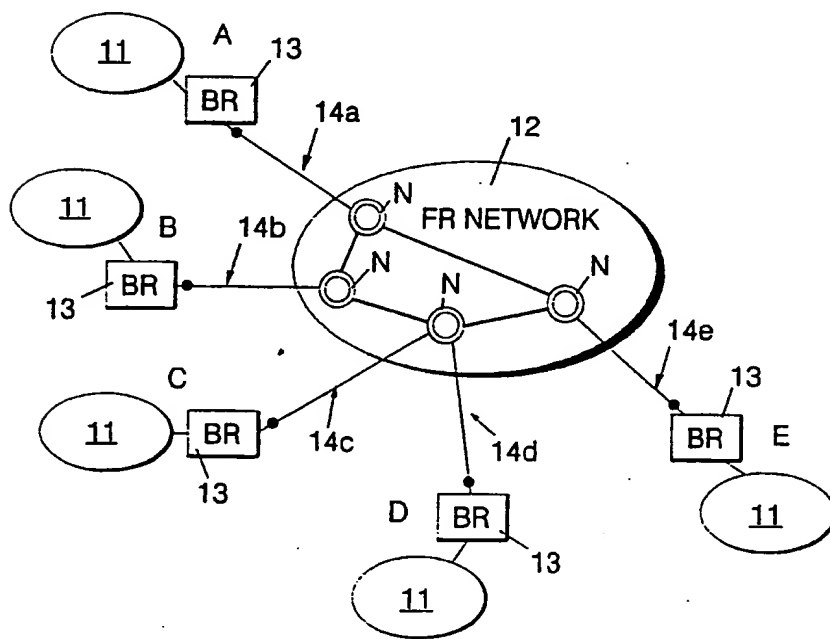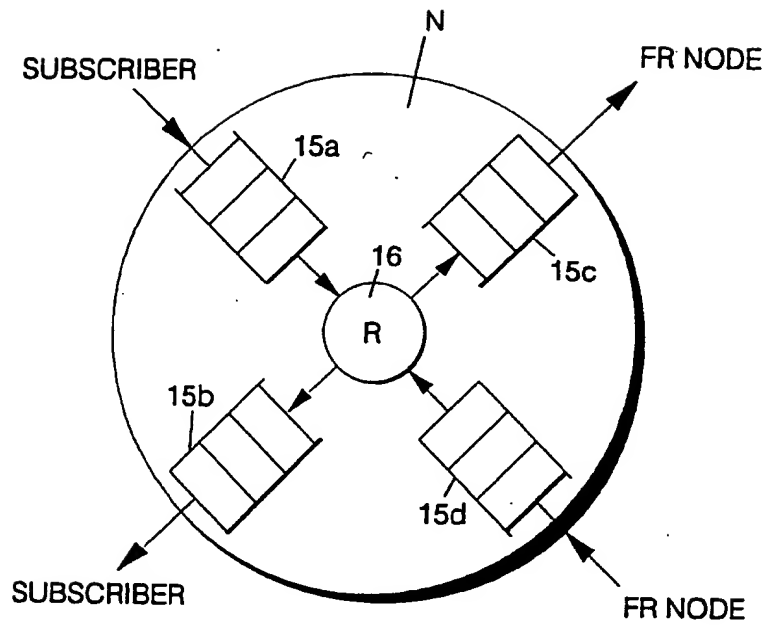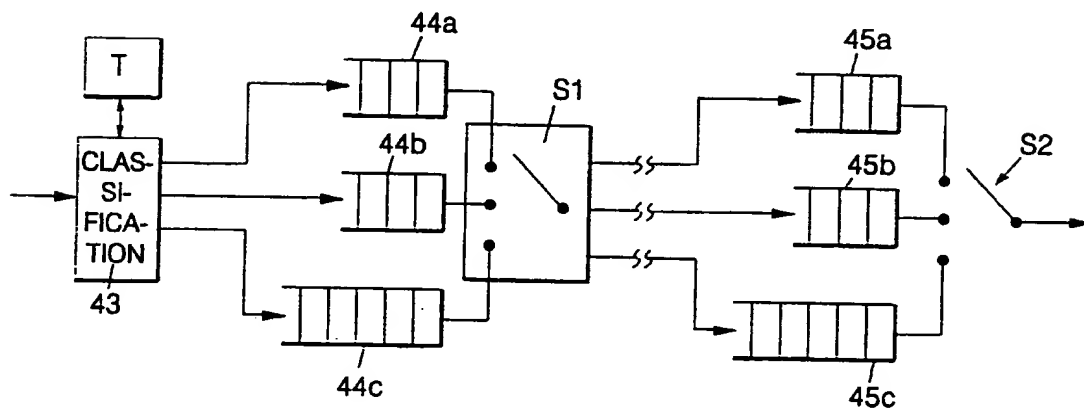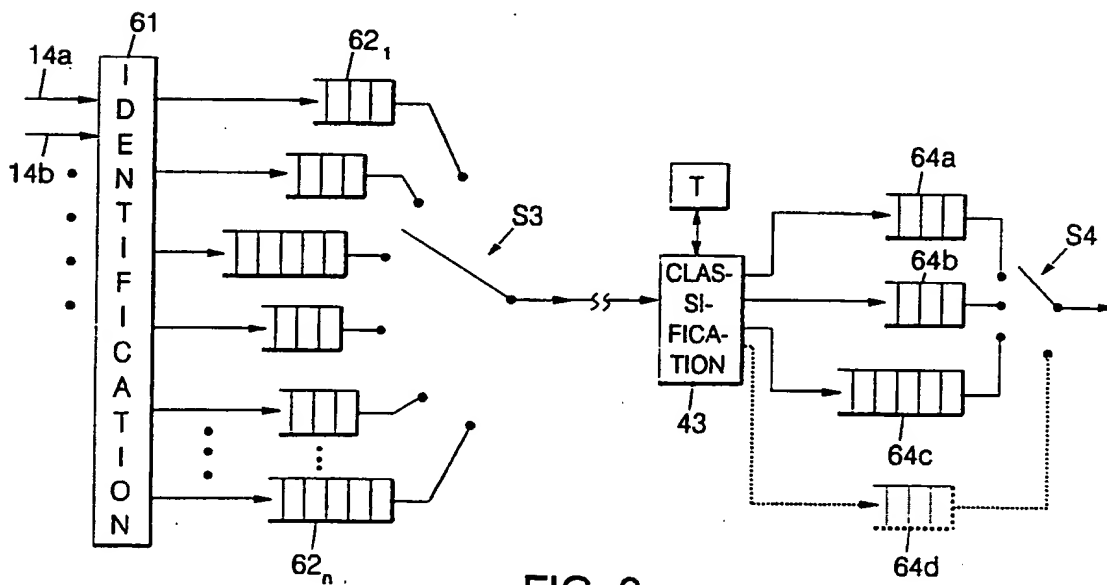
FIG. 1

FIG. 2

FIG. 3

FIG. 4

FIG. 5

FIG. 6
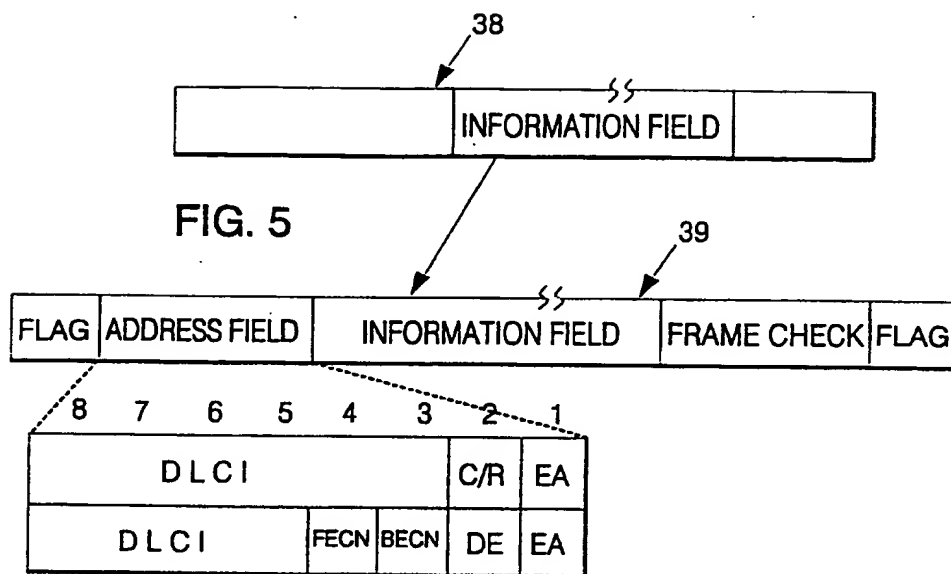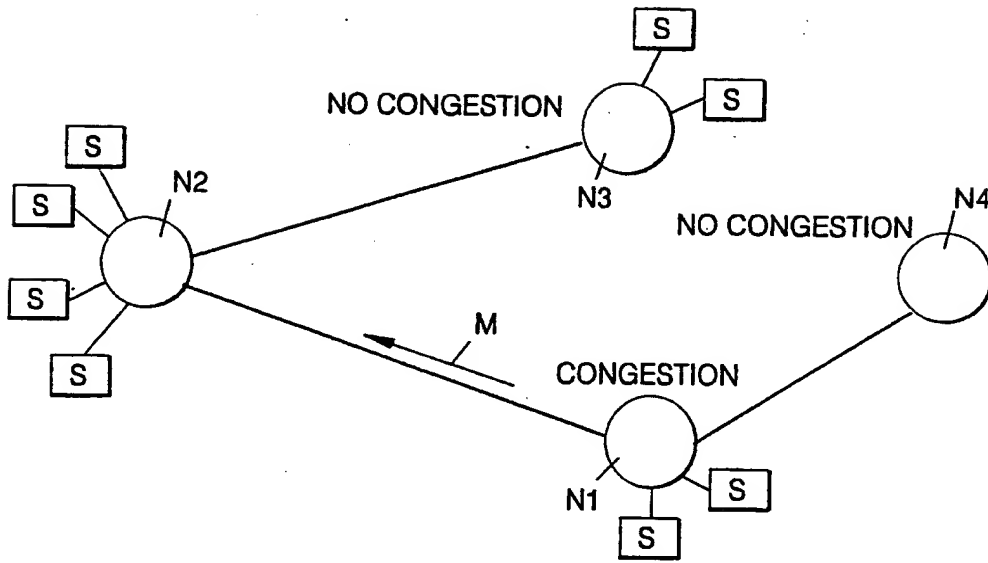
FIG. 7
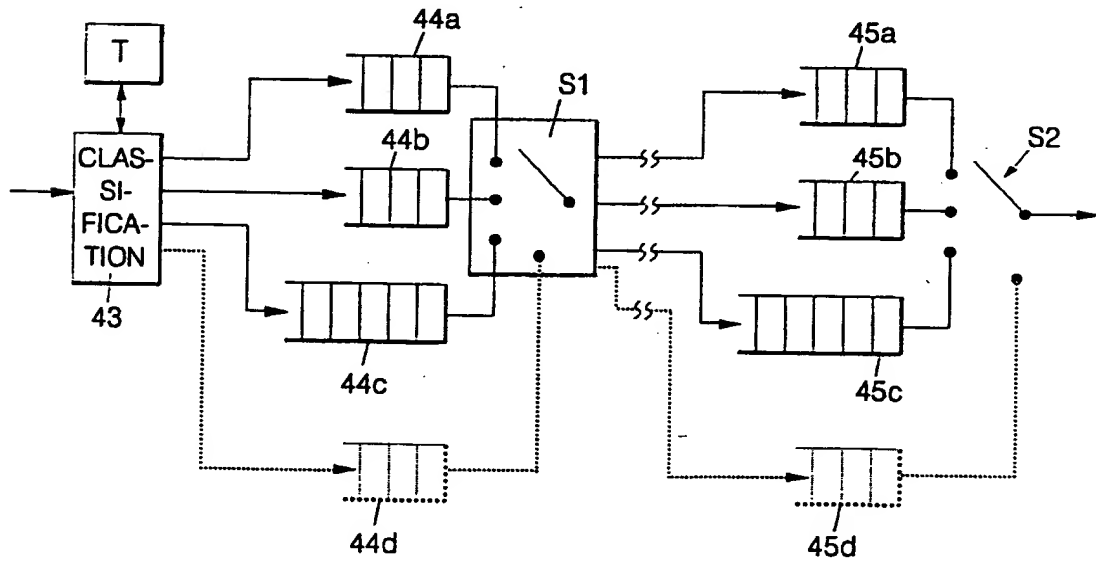
FIG. 8

INTERNATIONAL SEARCH REPORT

	International application No.
	PCT/FI 93/00537

A. CLASSIFICATION OF SUBJECT MATTER

IPC5: H04L 12/56

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC5: H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GB, A, 2261798 (DOWTY COMMUNICATIONS LIMITED), 26 May 1993 (26.05.93) --	1-9
A	US, A, 5014265 (ELLEN L. HAHNE ET AL), 7 May 1991 (07.05.91) --	1-9
A	US, A, 5159591 (SHINOBU GOHARA ET AL), 27 October 1992 (27.10.92) --	1-9
A	EP, A1, 0422717 (KONINKLIJKE PTT NEDERLAND N.V.), 17 April 1991 (17.04.91) -- --------	1-9

☐ Further documents are listed in the continuation of Box C.　　☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 March 1994	07 -14- 1994
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Nils Ekström Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

International application No.
PCT/FI 93/00537

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
GB-A- 2261798	26/05/93	EP-A- 0544452	02/06/93
US-A- 5014265	07/05/91	CA-A- 2029054 CA-A- 2030349 EP-A- 0430570	31/05/91 31/05/91 05/06/91
US-A- 5159591	27/10/92	CA-A- 2023446 EP-A- 0436069	06/06/91 10/07/91
EP-A1- 0422717	17/04/91	CA-A- 2026872 JP-A- 3145243 NL-A- 8902504 US-A- 5224091	10/04/91 20/06/91 01/05/91 29/06/93

Form PCT/ISA/210 (patent family annex) (July 1992)